PB87-147203

Use of Spatial Patterns and Their Uncertainty Estimates in the Model Evaluation Process

(U.S.) Environmental Protection Agency
Research Triangle Park, NC

Jan 87

U.S. Department of Commerce
National Technical Information Service
NTIS

PB87-147203

EPA/600/D-87/031
January 1987

THE USE OF SPATIAL PATTERNS AND THEIR UNCERTAINTY ESTIMATES IN THE MODEL EVALUATION PROCESS

by

Robin L. Dennis*
Atmospheric Sciences Research Laboratory
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

Steven K. Seikop
Program Resources, Inc.
search Triangle Park, NC

EPA Project Officer
John F. Clarke

*On assignment from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce.

ATMOSPHERIC SCIENCES RESEARCH LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
RESEARCH TRIANGLE PARK, NC 27711

REPRODUCED BY
U.S. DEPARTMENT OF COMMERCE
NATIONAL TECHNICAL
INFORMATION SERVICE
SPRINGFIELD, VA. 22161

TECHNICAL REPORT DATA *(Please read Instructions on the reverse before completing)*		
1. REPORT NO. EPA/600/D-87/031	2.	3. RECIPIENT'S ACCESSION NO. PB87 147203/AS
4. TITLE AND SUBTITLE THE USE OF SPATIAL PATTERNS AND THEIR UNCERTAINTY ESTIMATES IN THE MODEL EVALUATION PROCESS		5. REPORT DATE January 1987 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Robin L. Dennis and Steven K. Seilkop[1]		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS ASRL U.S. EPA RTP, NC 27711 [1]Program Resources, Inc. Research Triangle Park, NC 27711		10. PROGRAM ELEMENT NO. N104/C/05/05-6069 (FY-87) 11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS Atmospheric Sciences Research Laboratory - RTP, NC Office of Research and Development U.S. Environmental Protection Agency Research Triangle Park, North Carolina 27711		13. TYPE OF REPORT AND PERIOD COVERED 14. SPONSORING AGENCY CODE EPA/600/09
15. SUPPLEMENTARY NOTES		

16. ABSTRACT

Regional model evaluation is a new arena in which traditional residuals analysis has too little to offer. New techniques are needed that address quantification of regional pattern comparison, accounting for uncertainty and spatial correlation. Some of the more promising techniques are discussed. One technique, kriging, reasonably addresses the uncertainty in regional patterns and provides for an objective comparison. Other techniques, while useful, require that the comparison be based on subjective judgments. More work is needed in this area of regional pattern comparison.

17. KEY WORDS AND DOCUMENT ANALYSIS		
a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
18. DISTRIBUTION STATEMENT RELEASE TO PUBLIC	19. SECURITY CLASS *(This Report)* UNCLASSIFIED	21. NO. OF PAGES 8
	20. SECURITY CLASS *(This page)* UNCLASSIFIED	22. PRICE

EPA Form 2220-1 (Rev. 4-77) PREVIOUS EDITION IS OBSOLETE

NOTICE

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

THE USE OF SPATIAL PATTERNS AND THEIR UNCERTAINTY ESTIMATES IN THE MODEL EVALUATION PROCESS

Robin L. Dennis*
Steven K. Seilkop

Meteorology and Assessment Division
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711
and
Program Resources, Inc.
Research Triangle Park, NC 27711

1. REGIONAL MODEL EVALUATION AS A NEW ENDEAVOR

To evaluate the performance of air pollution models, it is essential to compare model outputs to actual monitoring data. The evaluation of regional models (scales of ~1000 km) poses new and important problems: (1) The domain of interest is regional, rather than local. Thus, a model's ability to identify several large-scale features is much more important than correctly predicting the concentration of a single, locally dominant feature. (2) The comparison of the spatial patterns of model predictions with those derived from observed pollutant fields takes precedence over comparisons at individual monitoring sites. In assessing the similarity of these patterns, it is important to consider not only their magnitude and location, but also their shape and positioning. A model's ability to replicate such pattern features may represent the presence or absence of important physical processes.

In the evaluation process, comparisons between observed and predicted patterns should be based on relevant and commensurate scales of spatial detail. The type of model under evaluation will dictate the appropriate scale. For example, the degree of filtering or smoothing of the physical processes in the regional models is greatly increased relative to urban scale models. Whereas urban models generally resolve the more important local variations, those same variations represent a perturbation on the larger, regional field. The scientific community must ascertain if the filtering of local perturbations has importance for a model's predictions. Model evaluation is concerned with the appropriate identification of similarities and/or differences between observations and predicted spatial fields. One must interpret the importance or meaning of such differences for an evaluation to be useful. Some differences may simply be due to the model's inability to resolve smaller-scale features, as shown in Figure 1 for a point source. In this case, the model should not be penalized for failing to mimic features that it is not intended to reproduce. On the other hand, some differences may imply that the model is lacking an important aspect of the physical/chemical system, as could be the case in Figure 2. In this example, the comparison of observed annual wet acidic deposition patterns of sulfate with predictions from a Lagrangian model shows regional scale differences, particularly with respect to shape.

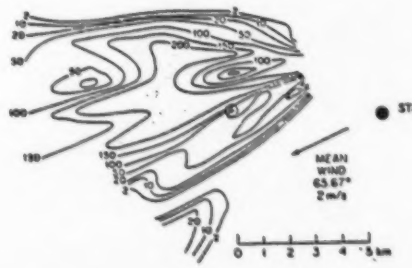

Fig. 1. "Coherent" plume footprint of one-hour concentrations of SF at 1300-1400 CST during EPRI Kincaid field experiment.

From our experience three aspects of regional observed and predicted data fields need to be considered when carrying out comparisons of model predictions with observations. (a) Subgrid scale effects. Local influences of emission sources and meteorology can affect the measurements at a particular monitoring site. Such local influences perturb the local measurement away from what would be the smoother, more regionally representative pollutant field. Subgrid scale features introduce variability about the regional pattern making the determination of the regional pattern less certain. (b) Measurement error. Random error in the measurement process, assuming there is no bias, introduces uncertainty into the determination of the magnitude of a regional pollutant field at a particular location. This uncertainty further exacerbates the problem of determining a regionally representative spatial field. (c) Spatial correlation. There is nearly always a fair degree of spatial correlation in regional pollutant fields. If the correlation is not taken into account, estimates of uncertainty associated with filtering the observed regional pollution field will be inaccurate.

2. APPROACHES TO ESTIMATING REGIONALLY REPRESENTATIVE POLLUTANT FIELDS

Because of the uncertainty in individual site data, a reasonable approach for evaluating model

* On assignment from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce.

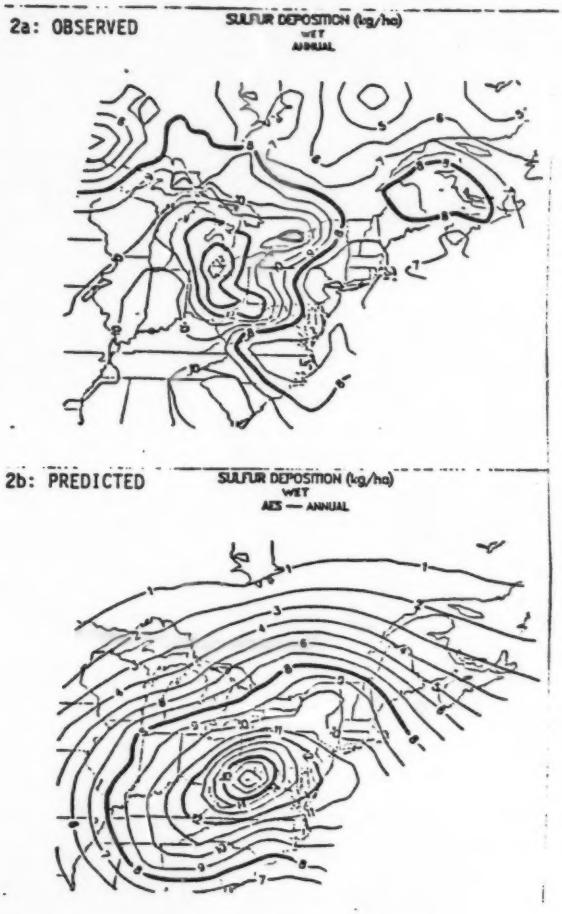

Fig. 2. Regional pattern of sulfur wet deposition for 1980 across the Northeast.

performance is to compare spatial patterns derived from model outputs and observed data. If appropriately estimated, these patterns retain the general spatial features of the data by smoothing through random measurement error and local spatial variation. However, these sources of variability induce uncertainty into the estimation of spatial patterns. In addition, spatial correlation needs to be taken into account in determining the degree of uncertainty associated with the estimation.

Traditional interpolation techniques are often used to smooth spatial data. However, they are in general deterministic, and do not use data weighting schemes that utilize the degree of variability in the data. Most importantly, these methods have no way to ascertain the accuracy of the characterizations that they produce. Statistical smoothing procedures redress these deficiencies. We have examined several such techniques for comparing spatial pollutant fields. In particular we will discuss the use of regression and of a statistical interpolation method known as kriging, both of which provide uncertainty estimates about an estimated surface. We will also discuss techniques that provide an understanding of differences in positioning of patterns.

Least squares regression is a standard method which might be considered for estimating spatial surfaces. Unfortunately, it has several undesirable characteristics. First, unless an extremely complex regression model is used, the method oversmooths the data and loses a great deal of spatial detail (see Figure 3 and compare it to Figure 2a). Oversmoothing casts doubt on the variance estimates that are produced, significantly inflating them. In addition, it is difficult to justify the critical assumption of spatial independence of regression errors. Failure to meet this assumption also results in the inaccurate estimation of variances. A positive distance-related correlation can result in serious underestimates of variance and inappropriately narrow confidence intervals for the estimated pattern. Because of such serious compensating errors, the user is unable to assess the reliability of error estimates obtained via least squares regression.

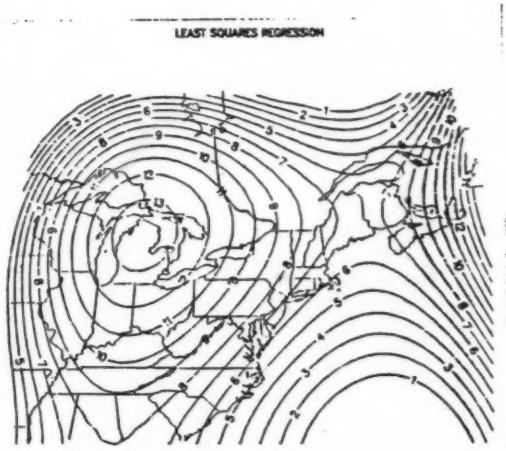

Fig. 3. Regression fit of observed annual sulfur wet deposition for 1980.

Weighted least squares is a potential alternative to address the correlated nature of the data. However, its results would also be questionable due to the instability of the variance- covariance matrix estimated from typically small amounts of available data (<100 sites). It would also suffer from the problem of oversmoothing. With such serious drawbacks we do not recommend use of standard regression techniques as a means to estimate spatial patterns.

The family of geostatistical interpolation methods, known as kriging (Delhomme, 1978), provides an attractive alternative to standard statistical methods for estimating surfaces. Kriging departs from traditional deterministic interpolation methods in that the determination of interpolation weights depends on the covariance structure of the regionalized

variable. A basic and critical part of kriging is in characterizing this structure through a function known as the semi-variogram. It is defined by:

$$G(h) = Var - Cov(Zi,Zj)$$

where G is the semi-variogram function; Zi and Zj are measurements at points that are separated by a distance of h units; Var = variance of the regionalized variable; Cov = covariance; and h = distance. The semi-variogram is estimated from the sum of the squared differences of all sample data values (see Delhomme, 1978 for details). It is used in combination with the relative geometry of interpolating points to derive interpolation weights. The estimator that results from this procedure has several desirable characteristics. First, it is an exact interpolator; the interpolated value at a point of measurement equals the measured value. This is not true of all interpolation procedures, e.g. least squares. Second, the kriging estimator is also the best linear unbiased estimator. This means the kriging estimator has the smallest estimation variance among unbiased estimators that are linear functions of the data, assuming the semi-variogram is correctly estimated.

"Simple" kriging is dependent upon the assumption that the underlying process is stationary. However, if the variable contains a strong spatial trend component, simple kriging will provide an overestimate of interpolation variance. To redress the nonstationarity problem and account for trend, "universal" kriging has been developed. Polynomial trend terms are introduced into the system of equations used to determine optimal interpolation weights. However, this technique relies on correctly predefining the polynomial form of trend and using a semi-variogram which does not include the trend component in its estimation. Unfortunately, the estimation of trend-free variograms has not been satisfactorily resolved and the necessity of having to preselect trend forms is a definite drawback. There is an alternative to these techniques known as generalized covariance kriging, but it requires extremely complex computing algorithms and equidistant points.

Because of the difficulties in applying generalized covariance kriging, we have concentrated on simple and universal kriging. If applied appropriately, simple kriging can be used as an alternative to universal kriging with its associated problems (Journel and Huijbregts, 1978). In fact, we have found that universal kriging is in some ways inferior to the basic technique. For example, a simulation study (Seilkop, 1983) showed that failure to eliminate trend from the variogram affects universal kriging's interpolation variances more adversely than those of simple kriging. Using actual data fields and cross validation techniques, the interpolation error estimates from universal kriging have also been found to be inappropriately large. For regional fields with moderate trend, as is the case for acidic deposition data, we have found that the bias in the estimated interpolation variance from simple kriging is on the order of 15-20%. In addition, we have found that strong gradients over short distances have only a minor effect on the estimates of error variance. This latter insensitivity is primarily due to the relatively large contribution that measurement error makes in the estimation of semi-variograms.

At this time, we recommend the use of simple kriging on regional fields, acknowledging that the error variances so produced will be too large. However, the bias is not large and its direction is known. As a consequence, comparisons that rely on the interpolation variance estimates will be less stringent than they should be.

3. COMPARING SPATIAL FIELDS USING KRIGING

Kriging was used in the International Sulfur Deposition Model Evaluation (ISDME) to develop spatial patterns with uncertainty envelopes for comparison of observed data with model predictions. Eleven statistical, Lagrangian and linear chemistry Eulerian models were evaluated. The comparisons of regional deposition fields were extremely useful for model evaluation. Both the model output and observed data were interpolated using kriging to 1/2 degree by 1/2 degree grid points in order to develop spatially commensurate regional patterns for comparison. An example for each is shown in Figure 2. The interpolated patterns from model predictions were developed two ways as a check: first, from the predictions at the monitoring sites of the observations and second, from model-gridded output. The two kriged regional patterns were essentially identical.

Based on the error variance estimates from simple kriging, the two-standard deviation uncertainty envelope about the observed field was computed. The magnitude of this uncertainty band, in relative terms, was approximately +/-50% over the areas of major deposition in the northeastern U.S and southern Canada. The two-standard deviations value is used to approximate a 95% confidence interval. Measurement error accounts for roughly 1/4 to 1/3 of the uncertainty. The estimate of measurement error based on the semivariogram fit to the observed data was consistent with estimates based on co-located sampling data. The typical uncertainty about the model output pattern for Lagrangian and statistical models is of the order of 1/10th that of the observational uncertainty. The factor of ten difference in variance indicates the higher degree of smoothing produced by the models compared to the observations. Thus, the observational uncertainty dominates a statistical comparison of difference between regional patterns of observed and predicted acidic deposition which incorporates both model and observed regional field error variances.

Given the uncertainty envelope, the regions of difference greater than a chosen threshold can be displayed. Figure 4 shows a map of the areas where the model predictions are greater than two-standard deviations from the observed data. Differences are quantified by defining subregions of the map and computing the percent of area of the subregion falling outside the threshold. This procedure provides a comparison of model predictions with observed data that incorporates the uncertainty

in the monitoring data as well as the uncertainty associated with interpolating model predictions. In addition, it provides a clear sense of regional model performance.

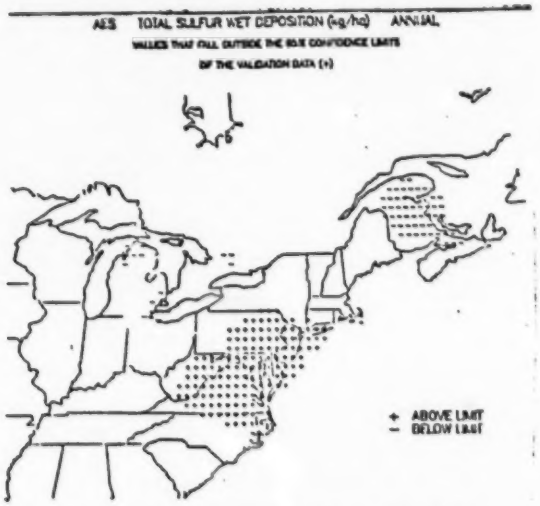
AES TOTAL SULFUR WET DEPOSITION (g/m2) ANNUAL
VALUES THAT FALL OUTSIDE THE 95% CONFIDENCE LIMITS
OF THE VALIDATION DATA (+)
+ ABOVE LIMIT
- BELOW LIMIT

Fig. 4. Comparison of observed and predicted patterns indicating differences across space.

4. ASSESSING SPECIFIC PATTERN DIFFE RENCES

The interpretation of pattern differences can be strongly influenced by misalignments between patterns, that is, differences due to translation and/or rotation. This is especially true for residuals analysis. Such misalignments complicate the determination and interpretation of bias. They also complicate the determination of shape differences via the comparison of spatial fields as presented above.

Three measures were examined for their ability to objectively quantify what the eye can see in graphical comparisons. The distance between centroids (centers of "mass") of observed and predicted fields provides a measure for translation. Defining the principal axes of a field via Principal Components Analysis (PCA) relative to the centroid provided a measure for rotation. Finally, shifting the pattern until the spatial correlation between the observed and predicted fields is maximized provides a measure of translation and/or rotation.

The centroids are influenced too greatly by the far reaches of a pattern in the case of acidic deposition fields where the ratio of maximum to minimum is less than three. Therefore, differences that the eye perceives are not captured by differences between centroids. Measuring the pattern shift by maximizing the spatial correlation yields a vector of translation much more in accordance with what the eye perceives (graphical analysis). A comparison of the results of the two measures is given in Table 1 for the 11 ISDME models.

The determination of major and minor axes via PCA, and hence differences in rotation of a pattern, is also influenced by points distant from the main pattern features. Thus, one must be careful when using this measure. The rotational shift estimated via PCA is similar to that estimated via the maximization of the spatial correlation under rotation. The spatial correlation measure will probably be the more reliable measure because we expect it to be able to better handle more complex pollutant fields such as surface ozone, in which there can be several dominant directional features.

	AES	AST	MOE	REL	RIV	RTM	SEA	SIM	STA	TAR	UMA
ANNUAL											
1. Centroids (km)	119	75	56	136	43	75	48	114	119	72	51
2. Shift (km)	397	333	220	467	311	348	330	347	326	311	346
SPRING											
1. Centroids (km)	51	68	46	62	30	25	16	72	94	46	32
2. Shift (km)	347	440	110	156	225	550	110	186	134	199	118
SUMMER											
1. Centroids (km)	100	63	51	147	33	54	10	127	42	44	42
2. Shift (km)	190	220	136	348	220	141	190	453	115	246	156
AUTUMN											
1. Centroids (km)	139	65	71	162	71	97	65	125	147	99	36
2. Shift (km)	663	440	330	592	347	934	246	454	592	440	453
WINTER											
1. Centroids (km)	146	43	82	144	46	112	42	71	151	32	67
2. Shift (km)	704	492	245	704	641	737	467	467	663	397	592

Table 1. Spatial shift of predicted pattern relative to observed for the two methods of centroids and spatial correlation.

Unfortunately, no estimate of the sample variation of the principal components is available. Thus, there is no way to quantitatively assess the similarity of components from observed and predicted fields. In principle, one could estimate the magnitude of sampling error associated with the principal components. However, this requires not only a simple perturbation of lbserved points by an estimate of measurement error, but a simulation of spatial variability. Since points are not independent, it is necessary to adequately reflect the degree of spatial correlation, and this is not a trivial task.

5. OBSERVATIONS

Regional model evaluation is a new arena in which traditional residuals analysis has too little to offer. New techniques are needed which address quantification of regional pattern comparison, accounting for uncertainty and spatial correlation. We have discussed some of the more promising techniques. One so far, kriging, reasonably addresses the uncertainty in regional patterns and provides for an objective comparison. Other methods, while useful, require that the comparison be based on subjective judgments. More work is needed in this area of regional pattern comparison.

6. REFERENCES

Delhomme, J.P. 1978. "Kriging in the hydrosciences," Advances in Water Resources Vol. 1, No. 5, pp 251-266.

Journel, A.G. and Huijbregts, Ch. J. 1978. Mining Geostatistics. Academic Press, New York. 600pp.

Seilkop, S.K. 1983. "A comparison of universal and simple kriging in the interpolation of values from nonstationary processes," Report to U.S. Environmental Protection Agency, (Contract #3D63679NASA).

DISCLAIMER- This is an extended abstract of a presentation and does not necessarily reflect EPA policy.